Chapter 1

Energy Challenges for ICT

Giorgos Fagas, John P. Gallagher,

Luca Gammaitoni and Douglas J. Paul

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/66678

Abstract

The energy consumption from the expanding use of information and communications technology (ICT) is unsustainable with present drivers, and it will impact heavily on the future climate change. However, ICT devices have the potential to contribute significantly to the reduction of CO_2 emission and enhance resource efficiency in other sectors, e.g., transportation (through intelligent transportation and advanced driver assistance systems and self-driving vehicles), heating (through smart building control), and manufacturing (through digital automation based on smart autonomous sensors). To address the energy sustainability of ICT and capture the full potential of ICT in resource efficiency, a multidisciplinary ICT-energy community needs to be brought together covering devices, microarchitectures, ultra large-scale integration (ULSI), high-performance computing (HPC), energy harvesting, energy storage, system design, embedded systems, efficient electronics, static analysis, and computation. In this chapter, we introduce challenges and opportunities in this emerging field and a common framework to strive towards energy-sustainable ICT.

Keywords: ICT, energy efficiency, energy sustainability, low power, embedded systems, smart sensors, high-performance computing, data centres, Internet of Things

1. Introduction

The reliance of society on the use of information and communications technology (ICT) devices and systems is ever increasing. From the proliferation of e-mail and electronic document exchange, social media and apps to the ready use of mobile devices (already in their fourth generation), data analytics, and advanced computing to solve big challenges, there has

been a transformative impact on society. However, the expanding ICT use requires increasing amounts of electricity to run and it implies fundamental transformations of energy that result in energy lost in the form of heat as explained in Chapter 2. Several models exist on the energy/electricity consumption of ICT, and some of them will be referenced below and in the remainder of the book. Nevertheless, a conservative estimation currently puts around 4% of all electricity consumption and over 2% of all CO_2 emissions as the result of ICT use. If entertainment, telephones, TV, and media that are now being translated onto ICT devices and systems are added, then these consumption numbers approximately double. In a recent study, the share of ICT global electricity usage by 2030 was estimated at 21% in a likely scenario and 51% in the worst-case [1].

By any account, the increasing energy consumption and the associated CO_2 emission of ICT devices strain the targets of low carbon, resource efficiency, and competitiveness of any modern circular economy. At present, all ICT roadmaps still use cost or performance as the main driver and improved energy management as a secondary issue, i.e., energy issues such as production, efficiency, and storage are considered only if necessary to achieve cost reduction or performance enhancement. However, if ICT is to become sustainable in terms of energy, then energy must be the key driver for all ICT devices and systems. Sustainable energy was defined by the United Nation's Brundtland Commission "Our Common Future" in 1987 [2] as requiring fuel or energy sources that have the following criteria: fuel is not significantly depleted by continuous use; no significant pollution or hazards to humans, ecology, or climate systems; no significant perpetuation of social injustice.

There are two main aims that must be achieved in order to meet this vision. The first is that the consumption of energy by all ICT devices and systems must be reduced. The second is that the use of sustainable energy and, in particular, renewable energy systems must be increased to power the majority of ICT. In fact, these aims represent also strategic conditions for the future development of ICT itself:

- **High-performance computing systems** are the ICT-enabling technology for advanced mathematical modelling and numerical simulations that play a key role in scientific discovery and technological innovation. If we want to foster the realization of the next generation of high-performance computing (HPC), we need to increase energy efficiency of computing. Exascale computers capable of reaching 10^{18} operations per second require a substantial decrease in the amount of energy dissipated into heat compared to present standards. There is a significant drive for energy efficiency in computing architectures, both for designing next-generation hardware, from the fundamental devices of information processing to data storage architecture and communication networks, and for developing software tools and algorithms to increase the efficient use of the hardware.
- **Smart autonomous sensor systems** for the so-called Internet of things (IoT) scenario. IoT foresees that an ever-increasing number of intelligent, mobile, sensing, and communicating devices will be dispersed into ordinary appliances and tools of common use. But most applications require an IoT device to be miniaturized, energy-efficient, and autonomous so that it is portable and self-sustaining. To achieve this, the amount of energy required by such devices needs to be significantly reduced and conventional power management needs to be replaced with energy-saving devices and other methods to regulate power

http://dx.doi.org/10.5772/66678

supply and demand. Emerging autonomous sensors need to maintain ultra-low power (ULP) duty cycles and incorporate an energy harvesting source, an energy storage device, and electronic circuits for power management, sensing, and communication into sub-cm scale systems.

- **Data centres** have become a critical ICT infrastructure owing to software as a service, mobile cloud applications, digital media streaming, and the expected growth of IoT. Data centre energy consumption is currently growing at a compound annual rate of over 10%. Power and thermal monitoring and control as well as recovery of waste heat play a key role in reducing consumption and economic costs. However, ever more opportunities exist towards a comprehensive integrated energy management system to enhance the energy and power management of data centres in conjunction with renewable energy generation and integration with their surrounding infrastructure.

To bridge the gap between the energy and power requirements in ICT and availability from energy harvesting/renewable sources, a multidisciplinary effort is required to address energy and power management issues across the layers of ICT systems. Several concepts for energy efficiency and sustainability are discussed in the forerunner of this book (to be referenced as Vol. 1) [3] and in the other Chapters. In Section 2, we discuss the energy sustainability of ICT, and in Section 3, we present challenges and opportunities in the framework of the ICT system stack.

2. The energy and power issue for ICT

Energy is the "capability of performing work." Power, being equal to the work divided by the time that it takes to do it, indicates how fast this work is done for a certain amount of energy or how much energy is consumed to achieve a task, e.g., a computational operation (in a specified time interval). In Chapter 2 of Vol. 1, the concept of energy and its relationship to power is discussed. In the next chapter of this book, the fundamentals of energy transformations (and losses in the form of heat) in information processing are discussed. An introduction to measuring energy consumption in computing is provided in Chapter 3. However, in anticipation to the discussion below and the general theme of the book on ICT-Energy issues, it should be understood that apart from reducing consumed energy there is also a balance to be struck with the ability to perform critical tasks. For the former, energy-efficient computing architectures could minimize energy-loss operations in information processing, data transfer and communications. On the other hand, taking the example of a wearable glucose sensor does require specified amounts of energy to be available at regular intervals so that monitored data are transmitted over the necessary distance via an antenna. Bearing these considerations in mind, the power required to operate current ICT systems ranges from the mW level for small autonomous sensor systems to tens of MWs for HPC systems. In between these power levels lie a large number of devices including embedded sensors, mobile phones, smartphones, tablets, personal computers, servers, and cloud computing storage systems. The annual sales of many of these consumer systems are now at the 100 million to 1 billion per annum and, without taking into account the energy required for production, every device consumes a certain amount of energy that results in the emission of CO_2.

The ubiquitous use of ICT systems has been driven by the continuous scaling of silicon chips. The original drivers for scaling came from improving the performance of computers, but as the size of transistors was reduced, so was the power consumption enabling many portable systems to be developed. **Figure 1** presents a summary of the scaling of silicon chips and demonstrates how the performance of computers has improved over time along with the number of transistors physically produced on each chip. However, as the number of transistors increased according to constant electric field scaling (Dennard's scaling rules [5]), the device sizes started to become so small for significant quantum effects such as tunnelling to kick in. While previously integrated circuit technology (based on complementary metal-oxide-semiconductor—CMOS) was dominated by dynamic power dissipation, leakage of currents due to quantum tunnelling started to become significant for advanced scaled devices. Fred Pollack from Intel first suggested the problems of continuing the scaling of the 1990s without changes to the architecture where the chip power density would scale to that of a nuclear reactor [6]. Indeed in 2006, Intel released a chip with a power density higher than the core of a nuclear pressurized water reactor (PWR). Chip design was changed to reduce this power density (**Figure 2(a)**), but an analysis of the absolute power indicates that with the increasing number of transistors, the total power of microprocessor units (MPUs) is still increasing over time despite the reduction in power density (**Figure 2(b)**). More worrying is the increase in peak power dissipation of low power, portable MPUs that are being driven by applications such as video streaming. This has led to architectures such as the ARM Big.Little [8], which during normal operation can allow low power operation, but when computationally intensive tasks must be undertaken, the peak power will increase significantly as required by the video streaming applications for compression and decompression.

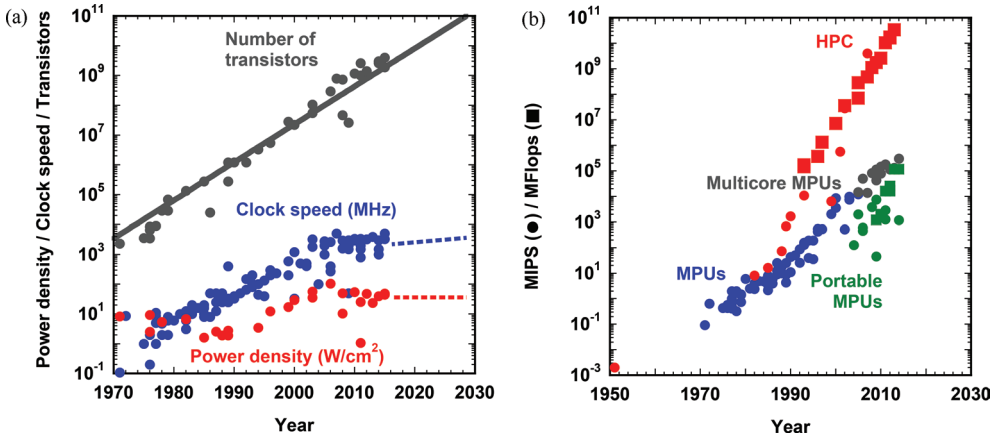

Figure 1. (a) The scaling of the number of transistors, chip clock speed, and power density as a function of time. (b) The performance of microprocessor units (MPUs) for computers, portable devices, and high-performance computing system as a function of time. The circles are data for million instructions per second (MIPS), while the squares are million floating point operations per second (MFlops) (Sources: Datasheets for processors from Intel, AMD, IBM, Digital, Motorola, Zilog, Samsung, Apple and Top 500 HPC [4]).

http://dx.doi.org/10.5772/66678

There are a large number of market surveys predicting the future of the ICT market and all of them suggest growth in a significant number of areas. **Figure 3(a)** shows the increase with time of the total number of ICT devices being sold each year. Only standard PCs and set top boxes are predicted to be static or decrease, while all other areas are predicted to grow significantly, suggesting that the number of ICT devices will continue to grow in the foreseeable future. As the number of ICT devices increases, and especially with the use of portable devices and the proliferation of the IoT, the amount of data being transmitted by the Internet and communication networks is also increasing significantly (**Figure 3(b)**).

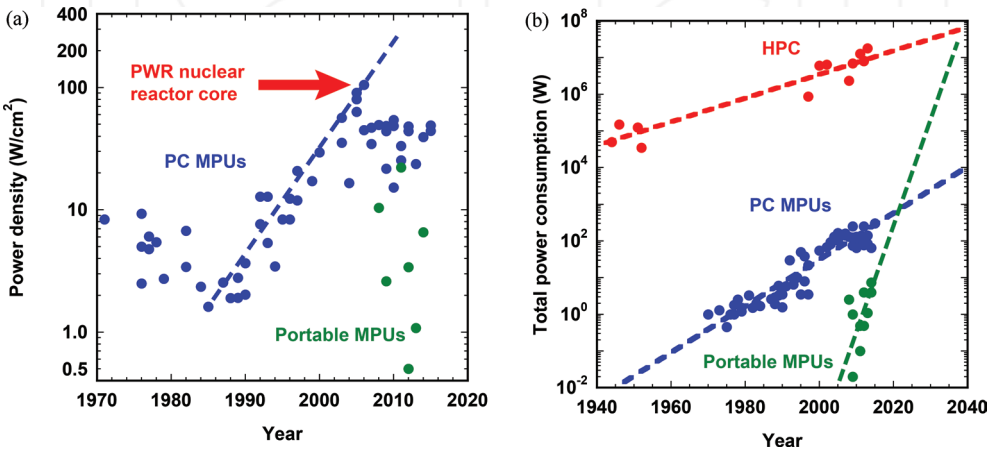

Figure 2. (a) The power density for PC and portable MPUs versus year of first release. The PWR nuclear reactor core power density is 102 W/cm² [7]. (b) The total power consumption of HPCs, MPUs for PCs (servers, desktop, and laptops), and MPUs portable devices (smartphones and tablets). Datasheets for the processors from Intel, AMD, IBM, Digital, Motorola, Zilog, Samsung, Apple, and Top 500 HPC [4].

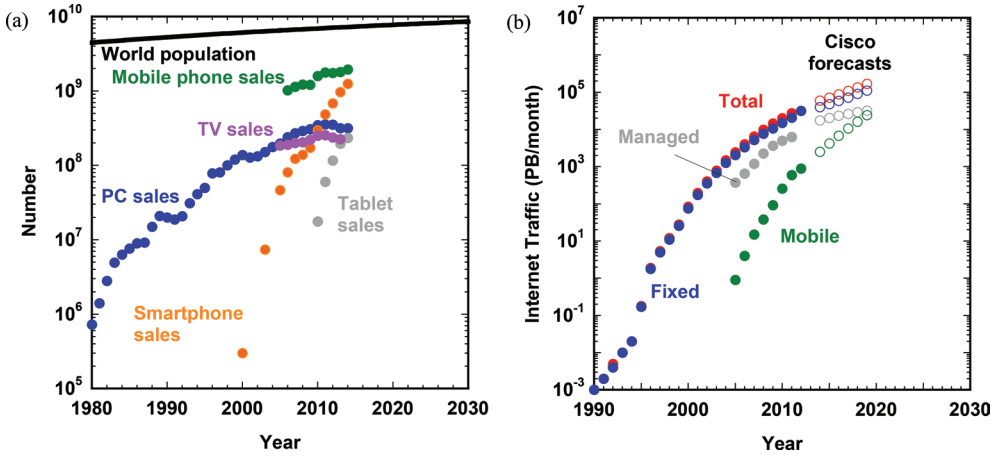

Figure 3. (a) The number of shipped end user device products per annum for personal computers (PCs—both desktops and laptops), mobile phones, smartphones, tablets, and TVs (Sources: [9, 10]). (b) The average Internet traffic in terabytes per month for each year (Source: [11]).

A number of studies have been looking at the energy consumption of ICT devices and systems (**Figure 4**). While several sources especially on web pages provide guesses of the total electricity consumption of ICT devices, there are a number of detailed studies that have tried to accurately estimate the total energy consumption and CO_2 emission [12–15]. These studies have used trade data to estimate the number of devices, analyzed average use and loading of devices and considered the power scaling to provide estimates of the total electricity consumption (**Figure 4**) and CO_2 emission (**Figure 5**). In particular, there are a significant number of studies investigating the energy consumption and scaling of the Internet as the present energy consumption of telecommunications is the fastest growing part of ICT energy consumption. The key message is that in 2015, around 4% of the electricity generated worldwide is consumed by ICT devices [16], which results in 1 billion tonnes CO_2 equivalent, that is, about 2.3% of the global emission of CO_2. These studies do not include TV and media uses of ICT systems or the CO_2 produced from the manufacture of the ICT devices and systems. The suggestion from reference [14] is that TV and media has 82% of the energy consumption of ICT but 131% of the CO_2 emission of ICT. As media is now being transferred to ICT devices and systems, a large part of this consumption may require being included in the total ICT consumption and emission in the future.

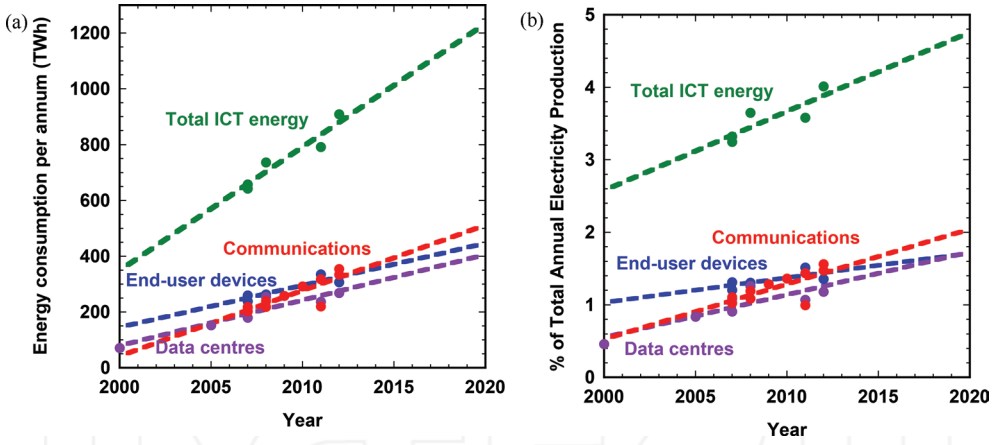

Figure 4. (a) The estimated energy consumption per annum for data centres, PC devices (including desktops, laptops, and tablets), communications (Internet, networks, mobiles, and smartphones), and data centres (servers including cloud computing) plus the total annual ICT energy consumption. Total ICT energy does not include any entertainment and media use such as TV, HiFi, DVD, CD, or radio. The ICT energy also excludes all manufacture and disposal of ICT devices (Sources: [12–15]).

An obvious way to reduce CO_2 emissions is to use sustainable and in particular renewable energy generation sources. **Figure 6** provides a comparison between the power requirements of ICT devices and systems and available sustainable energy generation technology. For a large-scale energy generation, the real issue is that most sustainable energy technologies require significantly higher capital outlay for installation (e.g., photovoltaic and hydro) and have long payback periods. Also many of the renewable sources cannot deliver a constant supply of energy and require both storage and/or alternative power supply mechanisms to maintain a

http://dx.doi.org/10.5772/66678

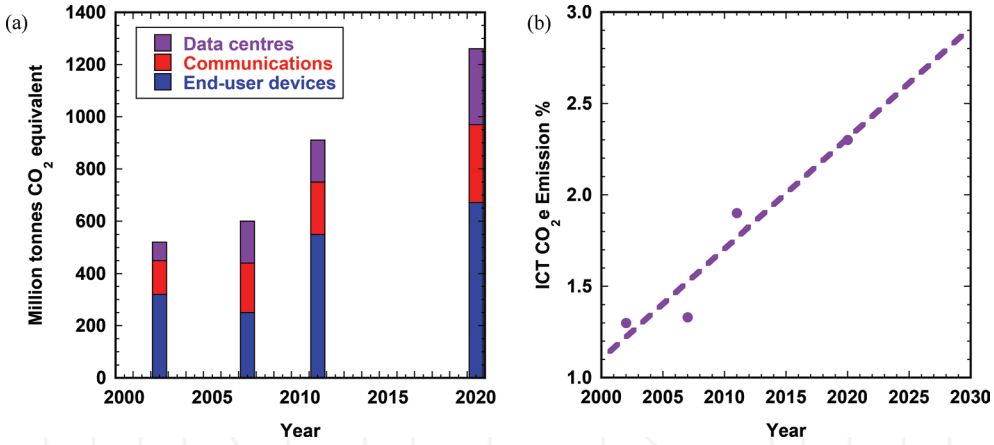

Figure 5. (a) The amount of CO_2 equivalent emitted from the manufacture and use of ICT equipment, infrastructure, and systems per annum along with predictions for 2020. (b) The percentage of ICT CO_2 equivalent emissions as a percentage of total CO_2 emissions (Sources: [14, 15]).

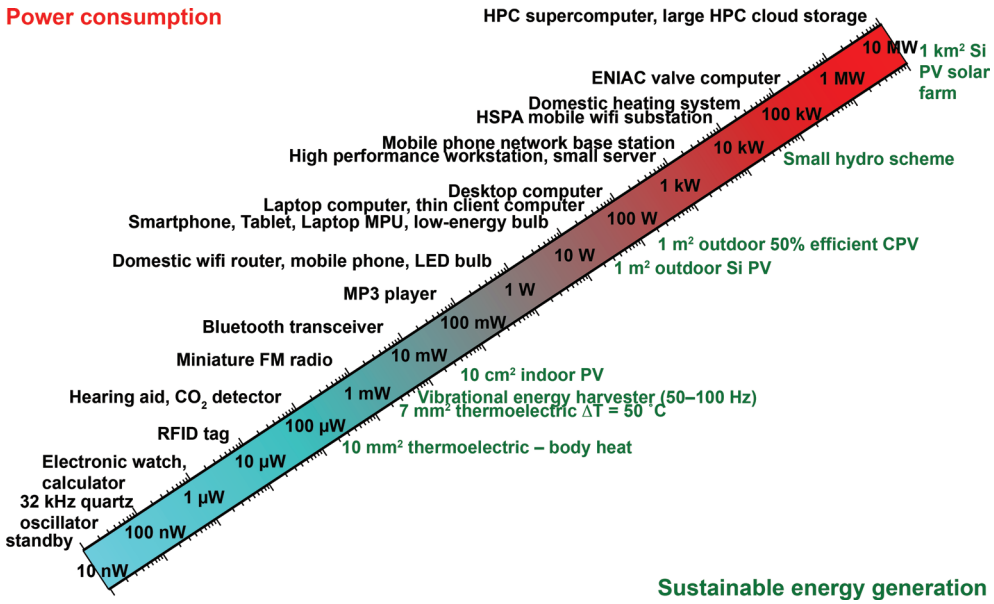

Figure 6. Typical power consumption of different ICT devices and systems versus the power that can be generated from sustainable/renewable energy generation devices and systems.

constant energy supply. At the small scale, batteries and super capacitors could potentially deal with such issues with advances in power management and disruptive technologies in micro-energy storage and harvesting. At the large scale, for HPC and cloud computing, the storage of

large amounts of energy is problematic and only pump-storage hydro can reach the required volume of energy in a sustainable manner. Such hydro schemes can only be built in suitable environments where large reservoirs with significant height difference can be built, and the location of such environments is seldom where the energy is required. Compressed air energy storage (CAES) is another suitable technology for bulk energy storage, but it requires underground caverns so it is also site-specific. An emerging technology that can address those issues is liquid air energy storage (LAES). LAES is modular and site agnostic, so it can be utilized for any size of energy storage and power rating up to several MWs. For specific applications, the advances and cost reductions in certain types of batteries (especially for flow redox and Li-Ion) also qualifies them as competitive technologies for significant energy storage.

3. Sustainable energy ICT: science/technology issues and opportunities

The system stack shown in **Figure 7** is a useful pictorial representation of the whole ICT system. While one might expect data centres (cloud) and HPC and smart autonomous sensors to be composed of completely different subsystems/layers, the system stacks are very similar in many areas and provide an opportunity to learn how to improve each from the experience of the other.

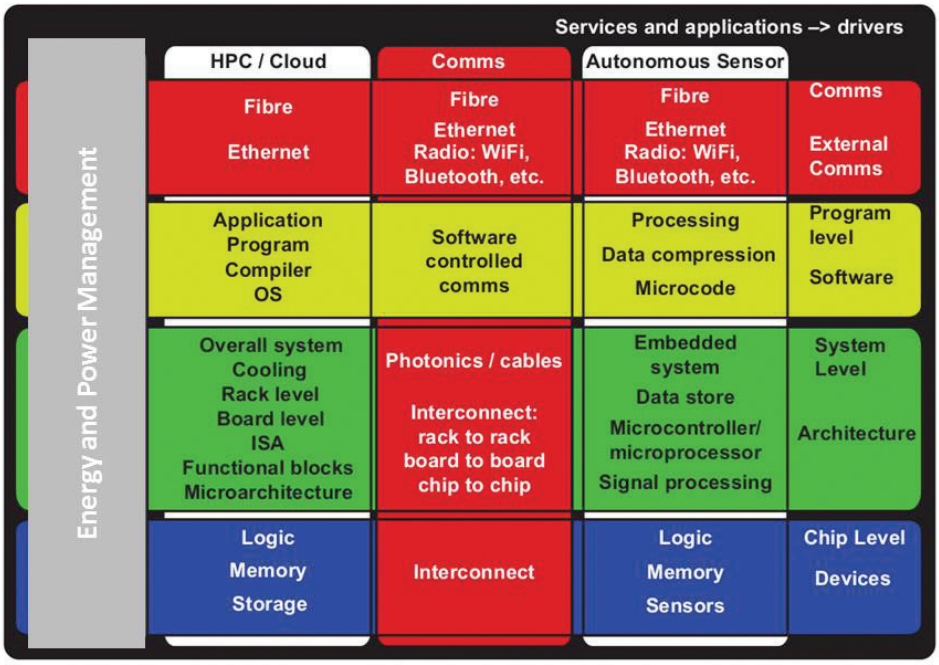

Figure 7. A schematic diagram of the system stacks for high-performance computing (HPC), cloud computing (Data Centre), and smart autonomous sensors. The figure indicates the common elements in the system stacks between these different ICT systems.

3.1. Chip level (devices)

From switches for logic and memory cells to transducing elements, devices provide the fundamental information processing. A generic ICT device can be viewed as a machine that processes information while transforming work into heat and heat into work. Pioneering research developed by J. Von Neumann and by R. Landauer in the last century has shown that information processing is intimately related to energy management [17]. An ICT device (see **Figure 8**) is a machine that inputs information and energy (under the form of work), processes both and outputs information and energy (under the form of heat). From this perspective energy dissipation via heat production and energy transformation processes are two aspects of the same topic: energy management at the micro- and nanoscales. For our purposes, energy efficiency is defined as the percentage of energy input to a device consumed in useful work and not wasted as heat. This definition, however, may not apply when we have to deal with processes taking place at the nanoscale (see Chapter 2 of Vol. 1).

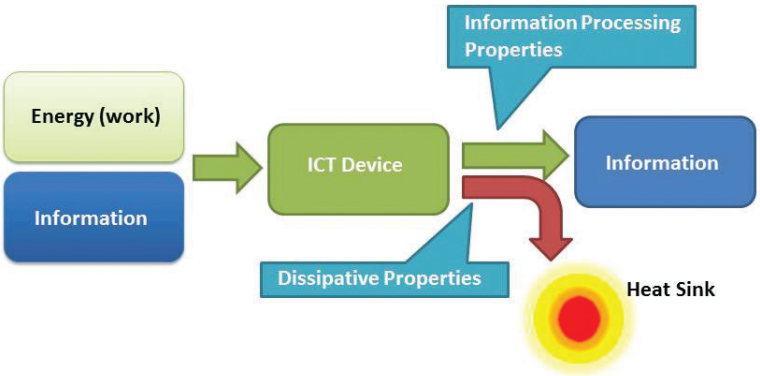

Figure 8. An ICT device is a machine that inputs information and energy (under the form of work) and processes both and outputs information and energy (mostly under the form of heat).

In sensor systems where conversion of external stimuli to signals is required, there exist several options of low-power transducer devices based on micro-/nanoelectromechanical systems (MEMS/NEMS) as well as optical and electrochemical sensing mechanisms. However, energy considerations become significant at the next stage that requires analyzing inputs and performing computational operations. As mentioned earlier, the side effect of the advances in the computation process is the increasing heat production. Here the workhorse has been the field effect transistor (FET), and in the last 40 years, the semiconductor industry has made impressive progresses in reducing the size of the CMOS components, thus increasing the computational density of microprocessors. New types of scaling rules as well as new designs and materials were introduced to reduce the energy dissipation following the breakdown of the Dennard scaling rules (where a number of fixed parameters in the transistor did not scale

in a standard linear, quadratic or cubic way with the gate length of the transistor, e.g., the voltage threshold at which the transistor is switched on into a digital "1" state).

While in principle, the switching energy of a metal–oxide–semiconductor field-effect transistor (MOSFET) could be reduced by reducing the supply voltage V_{DD} (indeed, this is what scaling over the last 40 years has been doing to reduce the energy and power dissipation), the finite bandgap of a semiconductor provides a lower limit below that the transistor will not switch on. Moving to another semiconductor with a smaller bandgap (than silicon) allows some improvement to reduce energy, but (apart from technological constraints) the fundamental physics of the p-n junction for the contacts provides another limit for switching MOSFETs on and off. A simplified analysis of a generic electronic switch at thermal equilibrium, modelled as a potential barrier separating two quantum wells (an idealized FET channel between source and drain contacts) showed that the channel could conceivably be scaled down to ~1.5 nm and the transistor could have a minimum switching speed of ~40 fs [18] (for the fundamentals of minimum energy of computing see also Chapter 7 of Vol. 1 and Chapter 2 of this volume). This is significantly smaller and faster than the FETs of today. FETs, however, have fundamental power and speed limits, and these cannot be overcome even by switching to new materials. For example, in order to avoid leakage at room temperature due to a finite height of the potential barrier, the voltage cannot be scaled as rapidly as the physical dimensions and the resulting power density for these switches at maximum packing density would be on the order of 1 MW/cm^2. The steep increase in CPU power density is mirrored by the increasing fraction of energy spent in cooling activities. Whereas there is still significant progress to be achieved by advanced CMOS, only by moving to a radically new technology can lower power dissipation potentially be achieved.

Figure 9 provides a comparison of present and future high-performance and low-power CMOS against future proposed device technologies. Present low power CMOS is already at 3 aJ per operation (excluding large fan-out and interconnect impedance) and only predicted to reduce by a factor of 3 over the next 13 years. The majority of improved future technologies do not produce any significant reduction in energy consumption per operation. Indeed a 100 times the Landauer thermodynamic limit appears to be difficult to better when all the new proposed device technologies are considered. This is 100 times the Landauer limit at 300 K and is related to the energy barriers in all the switching devices that provide a significant I_{on}/I_{off} ratio as required for the circuit architectures. A number of new types of devices suggest that this barrier can be circumvented. A switch with 8×10^{-20} W of switching power has been demonstrated albeit at the slow switching speed of ms [21]. More importantly, this switch demonstrates that the 300x Landauer limit can be broken.

Key device challenges can be summarized as:

- The cost of the lowest energy devices requires the latest CMOS technology node that now requires enormous economies of scale due to the cost of the foundries and technology.
- The scaling of transistors to smaller dimensions is now not expected to decrease the switching energy significantly.

http://dx.doi.org/10.5772/66678

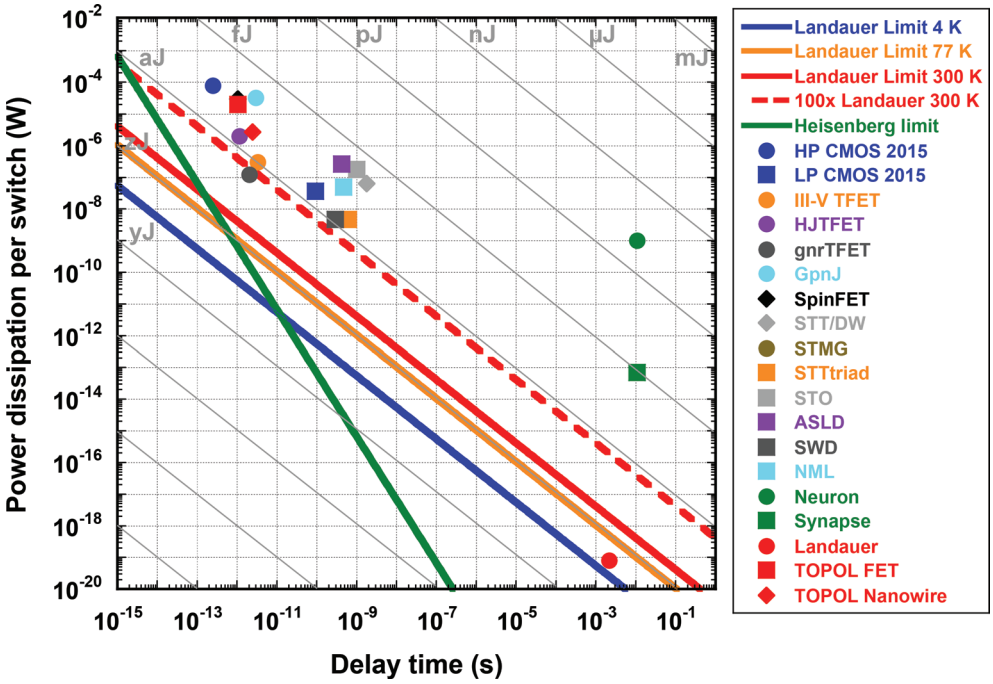

Figure 9. The intrinsic power dissipation per device switch versus delay time for CMOS and future devices as proposed in the ITRS future emerging technology roadmap [19, 20]. The diagonal grey lines are lines of constant energy. The figure indicates a limit of about 1 aJ per switch for CMOS devices and 250 zJ per device for the best proposed future switching device. Key: CMOS HP (high-performance CMOS), CMOS LP (low-power CMOS), iii–vTFET (III–V tunnel FET), HJFET (heterojunction FET), gnrFET (graphene nanoribbon TFET), GpnJ (graphene p-n junction), spinFET (Sughara-Tanaka spin FET), SST/DW (spin torque domain wall), STMG (spin torque majority device), STTtriad (spin torque triad), STO (spin torque oscillator), ASLD (all spin logic device), and NML (nanomagnetic logic). Results from the MINECC projects are the Landauer MEMS device (Landauer) [21], the Si CMOS FET from TOPOL (TOPOL FET) and the Si nanowire from TOPOL (TOPOL Nanowire).

- If there is a change of the basic switching device to move to a significantly lower energy technology, then circuit architectures, design tools, verification, operating systems, and software may require rewriting or complete changes for basic operation or optimal performance.
- Driving interconnects with multiple fan-out or an antenna have fundamental energy and noise limits that makes ultra-low energy consumption difficult.

Significant opportunities can be divided into continued scaling of conventional transistor devices (termed More Moore) and adding new device concepts for computation (termed beyond Moore or beyond CMOS) or new functionality onto the base CMOS technology (termed More than Moore). The key concept in More Moore is that the circuit architectures will be similar to present CMOS architectures and only the devices, voltages, or currents are changed. More Moore is looking at the challenges of scaling CMOS devices to dimensions below 10 nm,

and a significant portion of this work is now investigating new channel materials and device architectures such as gate-all-around nanowires that will allow higher performance with lower voltages for reduced power operation. The materials include Ge, III–Vs, semimetals, carbon-based materials, magnetic materials, and phase-change materials. Steep sub-threshold slope transistors allowing voltages for operation below the normal p-n junction limits are also being investigated. The major issue now being recognized by the semiconductor industry association and others is that scaling the transistor to smaller sizes may no longer result in lower energy devices; hence, More than Moore increased functionality is now a very diverse field of study:

- Devices with learning capability, e.g., devices that learn logic configurations and devices that learn by example.
- Integration of RF devices and components.
- Optical devices, e.g., Si photonics for communications and optical sensing modalities.
- Micro electro mechanical systems (MEMS) and nano electro mechanical systems (NEMS).
- Integration of sensors.

Beyond Moore is the opportunity to completely change how information is processed. This field is investigating the fundamental limits of information theory and if any of the known limits can be circumvented through innovative techniques. One example includes investigating if the Landauer limit requiring heat to be dissipated through the storage and erasure of information can be circumvented to allow zero energy switching. Spin wave devices are another example for a radical new technology that circumvents many of the limits of conventional transistor logic. Investigating methods of integrating device operation with energy harvesting has also been suggested where the heat dissipated and normally lost could be harvested to improve the overall thermodynamic system efficiency. The Beyond Moore research area could have a large impact in reducing energy consumption of ICT devices but is also the most difficult to implement into systems as solutions may be radically different from conventional CMOS technology, architecture, and systems.

3.2. System level (architecture)

3.2.1. Circuit microarchitecture

The microarchitecture level considers the integration of many of the underlying fundamental device technologies. Transistor technologies and silicon processes are combined into useful blocks such as memory (see, e.g., Chapter 4), control, and arithmetic that together form a computational device, often a processor or an accelerator. The scope for that integrated device is very broad, including ultralow power embedded devices, through general-purpose processors, up to high-performance network-on-chip components. As indicated in Section 3.1, current manufacturing of semiconductor devices has hit a fundamental efficiency limit called the "energy wall" that prevents reduction of energy consumption when transistor size scales down for forthcoming technology nodes. Both at a small-scale (embedded systems) and at a large-scale (HPC/Data centres), the so-called economic meltdown trend of Moore's law [22] transcends in a dramatic increase in the computation and cooling energy costs. Based on current projections,

http://dx.doi.org/10.5772/66678

a tenfold improvement in chip energy-efficiency is needed to maintain information technology (IT) energy scalability in the next decade. The ultimate limits from architecture designs are almost impossible to derive, but based on current technology, there is general agreement by academia and industry that new architectures are more promising to significantly reduce power consumption than improving the energy consumption of the basic switching device in the circuit.

The amount of energy consumption from a circuit architecture design for a given CMOS technology node is heavily dependent on how specific (i.e., optimized for a single or a few tasks) or how general (i.e., undertake many different computations) a design has to deliver. Applications specific integrated circuits (ASICs) designed for a single task can be optimized proving the lowest energy consumption, but such designs have no flexibility and cannot be reprogrammed. For microprocessors or microcontrollers that must be able to undertake a wide range of tasks, optimization to reduce energy consumption is significantly more difficult. Microarchitecture exploits what is physically possible with contemporary technology and presents an interface through which other hardware and software can use the device. In hardware terms, this interface is, of course, physical and will typically obey a specified protocol. In software terms, the processing device presents a set of possible operations through an instruction set architecture (ISA). If the ISA is a description of the behaviours of the device, then the microarchitecture is the implementation of those behaviours. Advances in physics, transistor design, and device manufacturing techniques can benefit microelectronic devices of all kinds; however, microarchitecture design decisions are heavily influenced by the target market of the resultant product. While all devices strive to achieve good efficiency, balancing performance and power consumption, the application area will dictate design constraints such as size, the energy budget, and maximum power. We may group microarchitecture characteristics grouped into four areas: deeply embedded, embedded/mobile, general purpose, and servers/high-performance. These are not necessarily strict boundaries and properties often transfer between areas over time, as technology or commercial pressures permit. At present, microarchitectures have significantly different drivers for HPC/data centres, general purpose (e.g., PCs), and embedded systems/portable systems.

Deeply embedded—An example of a deeply embedded device is the processor in a smart card. It must fit within a credit card form factor, cannot be modified once sent to the customer, be powered by a battery-backed device, and obey strict security protocols. A new generation of deeply embedded devices is smart autonomous sensors, which due to their ability to seamlessly integrate with the environment have given rise to cyber-physical systems (CPS) and IoT platforms. This requires integration of heterogeneous components dedicated to signal acquisition (e.g., analogue-to-digital converters (ADCs)), processing (digital signal processors), and environment manipulation (actuators). The embedded system may also typically include wireless communication and run on batteries together with energy harvesters. Hence, one or more of the following constraints may apply:

- Physical size must be small, from millimetres to a few centimetres, depending on application.
- The energy budget is finite as power may be intermittent or limited, often sub-watt or sub-milliwatt.

- Operating temperatures may vary significantly and the removal of excess heat quickly may not be possible.
- Reliability is essential, because servicing may be difficult, expensive, or impossible.
- Predictable behaviour may be required to guarantee safety criteria or always-correct device functionality.

All of the above points are relevant in an energy context. The energy consumption of the device dictates the temperature it runs at, the type of cooling required, how much processing and communication can be performed, and how long it will live. Predictable energy consumption is required to guarantee a particular battery life. To meet design goals within a small energy budget envelope, the design of each of the components is highly tailored to the targeted application (see e.g., [23, 24] and Chapter 8 of Vol. 1 for design consideration of a wireless sensor node). Circuit designs include low-voltage, power-efficient ADCs and filters, instrumentation amplifiers, domain-specific memory organizations, and schemes for energy/power management and transfer incl. energy-efficient passives.

The microarchitectures of deeply embedded systems take various forms, but the following traits are common:

- They provide predictable execution times for many or all of the ISA instructions they support.
- Their functional blocks, such as arithmetic and memory units, are often simpler than larger counterparts, to reduce power, improve predictability, and keep the device small.
- Their memory hierarchy is flat, avoiding caches that would impact predictability and increase device complexity.
- Programs may execute directly out of integrated flash storage, with RAM only used for read-write data.

Prolific architectures in this area include AVR, ARM's Cortex M-series, and PIC. They feature compact instruction sets (often 8- or 16-bit instructions), with short execution pipelines and in-order execution. This means opportunities for performance enhancement are limited, but their implementation is simple. The relative simplicity of such devices aids activities such as modelling their behaviour in order to predict energy consumption. However, properties such as the memory layout limit the types of application that can feasibly be run on such devices.

Looking forward, we can expect the safety and real-time requirements of deeply embedded systems to persist. However, research must strive to shrink these devices into sub-millimeter and beyond, with a desire for micro- or nano-watt power envelope devices with comparable performance today, while milli-watt devices deliver improved performance and capabilities. Multi-core is not yet common in deeply embedded systems, but in order to deliver the above improvements in the light of limits to Moore scaling, we can expect it to become essential. A particularly lucrative opportunity in deeply embedded research is a "zero power" idle or

http://dx.doi.org/10.5772/66678

sleep mode, with close to instant response time. When not responding to an event, the device should, ideally, consume no energy at all. However, following an event (such as inbound sensor data), it must be able to return to a responsive state.

Embedded/mobile—The line between embedded and deeply embedded is often blurry, but for the present purpose, we group *regular* embedded devices with mobile devices, in order to set the grouping in terms of energy requirements. Such devices may still have real-time constraints, small size requirements, and sub-watt power envelopes. An example of an embedded device is the controller chip on a hard disk or a solid-state disk. They cannot be replaced, so their failure effectively renders the entire disk useless. The software running on them is difficult or undesirable to update in the interests of data security. They interact with components that have strict timing protocols, and missed deadlines will potentially result in data loss or corruption. They must fit within the form-factor of the device, e.g., the whole device may be as small as 12 × 16 mm (the smallest of the possible M.2 format of expansion devices used in laptops at the time of writing). However, the performance requirement and available power are higher than deeply embedded devices, hence, this may be considered embedded rather than deeply embedded.

Mobile and embedded systems typically have the following microarchitecture properties:

- Heavily integrated into system-on-chip (SoC), providing various peripherals and multicore computational capabilities in a single chip.
- Larger storage and memory capacity than deeply embedded, in the order of gigabytes in current devices.
- Cache hierarchies for better memory performance.
- Sub-watt power constraints.
- Must fit within relatively small form factors, such as a mobile phone.
- More complex application sets and usage patterns.

Mobile and modern embedded devices feature performance, i.e., competitive with general-purpose hardware from less than a decade previous. A contemporary smartphone has more compute power than a ten-year-old desktop PC and consumes significantly less energy. Much of this is thanks to lower-level improvements, as described by trends such as Moore's Law and Dennard Scaling. If one is to expect the same to be true in another decade, then we must counter abstraction inefficiencies as systems become more complex. For example, May's Law states that software efficiency reduces to counteract any improvement in hardware efficiency. At the same time, as devices become even more integrated into lifestyles of the consumers, both users and app developers must be given better visibility of how and why energy is consumed by their apps. This transparency will encourage accountability for embedded software energy efficiency and narrow the gap between the best efficiency a hardware platform is capable of, and the efficiency achieved when running a particular set of applications. Energy-aware software engineering is discussed in Chapter 5.

General purpose—Global use of computing devices is shifting to a more mobile-centric approach. As such, the features of mobile devices often compete with those in more traditional general-purpose devices such as the desktop PC. However, general-purpose computing is less constrained than mobile and embedded computing, and this is reflected in the microarchitecture:

- Power envelope of tens of watts.
- Multiple layers in the cache hierarchy (three-layer caches or more).
- Less tightly integrated, with separate physical components for RAM, peripheral devices, etc.

The processor architecture may be similar to mobile devices. For example, both x86 and ARM instruction sets are used in desktop and mobile computing products. However, general-purpose microarchitectures that use these instruction sets may be more complex, with increased pipeline length, more functional units, multithreading capabilities, and higher operating frequencies. The trends in device usage suggest that general-purpose computing will merge with mobile computing, backed by servers such as those providing cloud services. To that end, the general-purpose computing category may eventually disappear, rather embedded/mobile will become the defacto general-purpose computing platform. Thus, research into low-energy microarchitecture may create more benefits if it focuses on embedded and server-related areas.

Servers/high performance—The properties of high performance and server devices are similar to general purpose, but their form factor and tolerances are different.

- Multiple servers may occupy dense racks.
- Power envelopes may be higher than general purpose due to more aggressive cooling.
- Underlying architectures are similar to general purpose, typically at leading-edge, with additional resilience features such as error correction.

Power dissipation per server must be considered in order to adequately provision the electricity supply as well as to extract the waste heat. HPC computer architectures such as server processors and MPUs, work within power constraints in the order of tens or occasionally hundreds of watts. One of the most promising lines is the development of energy-efficient processing architectures building blocks that would significantly enhance the energy-proportionality of server processing power at the deep submicron era (i.e., beyond 28 nm technology nodes). The development of these building blocks will be achieved by using emerging technologies such as fully depleted silicon on insulator (FDSOI) [25] and gate-all-around nanowires [26], and integrated microfluidic cooling and power delivery [27]. This development would require modelling the power and thermal dissipations involved in the processing units, memory hierarchy, and the cooling and power delivery networks at the server level.

Mobile architectures such as ARM are beginning to encroach into the server space, with examples such as Cavium's ThunderX processors and APM's HeliX devices, both of which use the 64-bit variant of ARM for server-grate compute and infrastructure roles. The continued proliferation of multi-core supports this movement, and it is likely we will see lower power

http://dx.doi.org/10.5772/66678

per device, but packed at a higher density so that overall the power per rack is still a challenge [28]. Interconnect between these systems also becomes of particular research interest, as the cost of data movement between these devices does not scale with the reduction in energy per processing unit.

A key challenge at the microarchitecture level is that energy efficiency advancements must take the form of a synthesis of improvements at lower levels and respond to the evolving needs at higher levels. We must strive to provide predictable, transparent behaviour, not just functionally, but for properties such as energy consumption. As discussed later, the cost of data movement must also be more readily exposed and significant effort must be put into efficiently moving (or avoiding moving) data around a system as its contribution to energy consumption will continue to increase. This must be complemented by novel, intuitive methods for presenting the increasingly heterogeneous resources that form emerging multi-core systems across the various device classes so that higher levels of the system stack can fully understand and exploit the underlying hardware. Opportunities for improvement at the microarchitecture level are along three main directions:

- Increased heterogeneity, with specialized functional blocks for particular tasks.
- Many-core systems, with various heterogeneous blocks as described above, combined with groups of homogeneous clusters.
- Network-based interconnects or complex, multilayered buses, to connect these many components, along with caches, memory, and peripherals, together.

All these pose scalability and programmability problems, some of which must be addressed at the microarchitecture level, while others can be dealt with in other levels in the stack.

3.2.2. System architecture

Similarly to the scenario of circuit architecture, system architecture will require multidisciplinary research efforts to achieve holistic energy-efficient design and management. Advances in computer architecture must be carried out in two information propagation directions: The first direction is bottom-up where technology-related parameters (e.g., FDSOI and Stacked-DRAMs) are propagated from new technologies and chip level to the large-scale data centre level. In particular, circuit-level technological parameters will be used in the server and data centre architecture explorations, while the chip-level cooling parameters will be exploited by the large-scale cooling and energy reuse technologies. The second direction is top-down where the target large-scale data centre operation characteristics (in terms of runtime workload variations at software level and infrastructure operating conditions), will be exploited to tune further the lower-level architectural, cooling, and technological design aspects.

In both mobile and HPC microprocessors, there are conflicting demands for high performance and energy efficiency. Circuit architectures to deliver these conflicting performance requirements are being addressed through the heterogeneous integration of a range of embedded cores. One example is the ARM big.LITTLE architecture [8], where smaller cores are employed to process simple, less demanding tasks to save energy, while larger cores are optimized for

the high-performance tasks which are more energy demanding. This is a general trend, and, in parallel to it, there are a wide range of complementary techniques that are either being used in commercial processors or are being researched in academia to reduce the power consumption through architecture design. These include the following:

- *Dynamic voltage and frequency scaling*—High voltages and frequencies are used for high-performance task delivery, while low voltages and frequencies deliver reduced energy consumption [29].
- *Clock gating and clock distribution*—This is a dynamic power management technique where additional logic switches are placed between the clock and the clock input of the processors logic to disconnect the logic preventing clock cycling when it is not required [30].
- *Power domains*—These are sections of a core in a processor that can be completely powered down to reduce energy consumption without removing the supply to the system [31].
- *Pipeline balancing*—This is the dynamic adjustment of the resources of the pipeline of a processor such that it retains performance while reducing the power consumption [32].
- *Caches and interconnects*—It has been demonstrated that too large caches waste energy while too small caches limit bandwidth and performance. Careful optimization is, therefore, required to both minimize energy consumption and maintain performance [33].
- *Dynamic partial reconfiguration*—A number of field programmable gate array (FPGA) manufacturers now offer dynamic partial reconfiguration that enables reconfiguration of parts of the FPGA while the other regions are still active [34].
- *Composable and partitionable architectures*—This is where a set of low-power small cores can be aggregated together dynamically to form a larger single-threaded processor when required for higher performance [35].
- *Coarse grained reconfigurable array architecture*—Designed to be reconfigurable at the module or block level rather than at the gate level in order to trade off flexibility for reduced reconfiguration time [36].

On a larger scale, some of the new architectural solutions that are being investigated in the context of data centres and HPC include the following:

Advanced cooling infrastructures and energy reuse

- A passive thermosiphon (gravity driven) cooling system for servers and racks, using energy only to remove the waste heat out of the data centre room.
- Innovative systems able to re-convert generated heat into electricity, such as the pressure reverse osmosis (PRO) process. The objective is to provide data centres a solution to absorb a part of their waste heat and reproduce electricity.

Data centre thermal control and management

- Resolution of complex runtime multi-objective optimization (performance, power, and temperature) in high-level IT workload management (allocation and scheduling) and

physical resource management (processors, memories, storage, network, and cooling infrastructure configuration).

- Exploration of several optimization schemes and control approaches such as adaptive dynamic programming, networked control, and predictive control at the large scale of data centre level.

System-level energy-efficient data centre architecture design

- Definition of more efficient system-level exploration approaches to develop energy-efficient thermal-aware architectures for servers, racks, and data centre rooms by using a holistic and tighter integration of computing and cooling power costs.
- Development of scalable simulation methods of large-scale computing systems that integrate power and thermal modelling to help data centre designers make robust predictions to anticipate potential failures of components and to accurately estimate the necessary provision in global energy during data centre conception and throughout its lifetime.

In a multiprocessor system-on-chip design (MPSoC), joint architecture optimization and integration of low-power components in novel architectures are the only way to keep increasing the performance while staying under the power budget. Solutions include application-specific and heterogeneous memory hierarchy, for instance, heterogeneous 3D architectures (stacked-DRAM) or efficient hardware implementation of components (analogue and digital) for ultra-low-power sensing. On top of this, all the on-chip components (both logic and memories) are increasingly affected by process variability, which means that they can no longer operate always under the best conditions, requiring self-adaptive architectures. Therefore, this scenario makes it more important than ever to develop scalable functional simulation frameworks to explore the limits of parallelization and global power reductions in MPSoCs. In the field of HPC and cloud computing, as volume server costs drop, electricity costs will emerge as a substantial fraction of the overall cost of ownership in HPC clusters and data centres. The fundamental question beyond the state-of-the-art is how to bridge the efficiency gap between emerging HPC, data-centric and scale-out workloads and modern server and network platforms. There is a large efficiency gap between existing server architectures and what the emerging data-centric scale-out workloads need.

Chapter 8 discuss architectural solution opportunities and implementation challenges in more detail.

3.3. Program level (software)

Software affects the amount of energy consumption in a system in a variety of ways. Possibly the most significant is that many facilities in a system operate at the request of software, meaning that processor cores and system-level consumers such as data communication or displays can consume more or less energy, depending on software algorithms. Accepted wisdom states that because of this, the best energy efficiency is achieved by having all software operate as quickly as possible, meaning that facilities can be disabled (minimizing their consumption) for longer periods. This is known as reducing static power. In addition, software also causes energy consumption through dynamic power, the cost of circuits charging and discharging as they signal digital information. Examples include the activation of different banks of memory

according to access patterns, driving of data buses, and switching activity as gates in arithmetic units converge on an output. While static power tends to be controlled by the algorithm behind a piece of software, dynamic power is governed by its particular implementation. Its relation to higher-level programming constructs tends to be poorly understood.

In need of disruptive solutions, we cannot rely on a hardware delivering better energy performance in the future and so the developer must contribute to energy reduction too. An obvious barrier at present is that very few software developers have much idea of how much power their programs dissipate or which parts of a program are energy hotspots. This might even be different for the same program from one platform to another. There are good software engineering reasons for the programmer's ignorance of energy, namely, to allow programs to be ported to different platforms and to allow program design as higher levels of abstraction. The large conceptual gap from hardware, where energy is consumed, to high-level languages and programming abstractions has been created by decades of computer science research and compiler advances. Somehow, the developer using a high-level language has to understand the energy induced by the software at the hardware level, without having to measure it on a machine. A clear theme emerges that the developers of the future will rely less on the performance of processors, but instead be energy-aware. This means that they tune their software to work optimally on the available hardware, or, if feasible, choose hardware specific to the software application.

Given a program to be implemented in some programming language and a hardware platform on which it is to be executed, we may ask whether it is energy-optimal. The energy limit for a software is a relative and pragmatic one; what is the least energy a given program should consume on a given hardware platform? We assume that in answering this question, the program could be redesigned to use a more energy-efficient algorithm (for that hardware platform). Software designers and developers typically first target functionality (getting the program to do what it is supposed to do), then performance (doing it as fast as possible), and third, minimization of code development costs (productivity). Optimization of performance is highly important in some fields, especially in HPC. This usually means optimizing for minimizing the time of a computation (i.e., optimizing performance and high speed). However, there has not been significant work on optimizing the energy consumption or understanding the ultimate energy consumption limits. Energy consumption of software is not subject to fundamental limits in the same sense as hardware. A software is written using programming languages and translated into codes that are executed by hardware. This provides significant opportunities for optimizing energy consumption. Large-scale integration (LSI) logic suggests that dedicated low-power hardware circuitry may save 20%, while changes to software to better control the power states could provide power savings of a factor of 3–5. In short, more energy is wasted by software than by hardware.

In a general-purpose piece of software, reducing energy consumption will mean reducing the amount of static and dynamic power the application draws. This process requires two distinct steps: that of identifying the sources of each kind of power, and then reducing that amount of consumption. Identifying static power consumption corresponds to evaluating the amount of time taken for the program to execute, while dynamic power requires either analysis of the machine instructions that the program compiles to, or some (more or less approximate) model

http://dx.doi.org/10.5772/66678

of the energy usage of higher-level software. Both of these techniques are complex, and not immediately available to modern software developers, limiting their ability to make decisions to reduce energy. The key research challenge is to bridge the conceptual gap and make energy consumption transparent through the layers. When programmers are energy-aware, there are a number of measures that can be taken to optimize for energy efficiency of software [37]:

- Choose the best algorithm for the problem at hand and make sure it fits well with the computational hardware. Failure to do this can lead to costs far exceeding the benefit of more localized power optimizations.
- Minimize memory size and expensive memory accesses through algorithm transformations, efficient mapping of data into memory, and optimal use of memory bandwidth, registers, and cache.
- Optimize the performance of the application, making maximum use of the available parallelism.
- Take advantage of hardware support power management.
- Select instructions, sequence them, and order operations in a way that minimizes switching in the CPU and datapath.

The translation and execution process is separate from the software and depends on tools such as compilers, interpreters, and schedulers and the hardware platform. All of these may influence its energy consumption and can be varied independently of the software itself.

Below are some of the opportunities to be explored regarding software execution:

- *Energy-aware algorithms*—Already today and increasingly important in the future, algorithms need to be able to respect power and energy constraints. Power and energy need to be added to the conventional design goals of performance and correctness, and metrics for assessment need to be established. Standard interfaces and application programming interfaces (APIs) for collecting power and energy information have to be developed, supported by accurate measurements through built-in hardware counters and sensors or external measurement devices.
- *Avoiding data transfer*—Data transfer at all levels including local memory to cache, within local memory, read/write to storage, and transfer through the network is expensive both in terms of time and energy compared to floating point operations. Algorithms need to be investigated that reduce the need for data transfer and which exploit and improve data locality.
- *Data compression*—The volume of transferred data can be reduced if the data is compressed. However, the compression itself is an additional computation that consumes time and energy. Algorithms need to be investigated for their feasibility to use data compression, and the trade-off with time and energy needs to be taken into account.
- *Multiple precision algorithms*—Not all applications require the full IEEE-754 double precision accuracy, which is however often used by default. Algorithms need to be investigated for their feasibility to use multiple, lower precision data formats. This might speed up the computation, reduce the memory requirements, and reduce the data transfer, which all can

contribute to a reduced time and energy consumption. However, the numerical properties must be carefully considered since multiple precision algorithms might experience different numerical stability.

- *Reducing synchronization*—In most algorithms, at some point, the computation must be synchronized across the machine that usually imposes waiting and idle times. One example of a global synchronization that appears in myriads of algorithms is the computation of the dot product, where all processors need to provide a local contribution and the aggregate result is distributed. Research is needed for restructuring of existing and development of new algorithms that reduce the synchronizations.
- *Randomization and sampling algorithms*—Another approach to reduce synchronization and data transfer is the use of randomized or sampling-based algorithms. Such algorithms work on decoupled, independent subparts or instances of the problem and synchronize only locally or occasionally. However, such algorithms may show nondeterministic behaviour, or even sometimes fail to return a result. Research should address both the modification of existing deterministic algorithms towards randomization and sampling, and the developments of new algorithms. The numerical properties and the suitability for specific applications need to be considered.
- *Adaption to load imbalance*—Ill-balanced codes can infer substantial penalties on performance and energy consumption. Load imbalance may occur even for initially well-balanced simulations due to different numerical properties of the problem evolving in different temporal or spatial domains, after recovery from failure, or imposed by energy management.
- *Scheduling and memory management*—In order to efficiently use the massively parallel and heterogeneous platforms, power and energy-aware runtimes, scheduling, and memory techniques are required.
- *Autotuning algorithms*—Complementing the scheduling and memory management on the runtime level, algorithms need to be able to detect and tune themselves to the architecture. Numerical software libraries need to become able to choose the most efficient variant of an algorithm for a particular hardware and a particular application in an automatic way.

The lower energy bound limitation stems from the generality of hardware. The main benefit of software is that it is reconfigurable: one can take a general-purpose processor and deploy many different applications to it, perform field updates, and otherwise alter behaviour without altering hardware. This necessitates that the processor is substantially more generalized than the application, to allow for reconfiguration. It must have a large enough set of behaviours (i.e., be Turing complete) to be programmed, as well as having sufficient performance and resources to meet application budgets for time and space. This leads to a significant capability gap between hardware and software. On the one hand, the hardware must have a high capability to allow its reconfiguration for different applications, but on the other hand, any particular application will only require a subset of those capabilities. Dynamic voltage and frequency scaling (DVFS) and gating techniques allow disabling un-needed capabilities to some extent, while heterogeneous systems can provide application-specific acceleration

facilities (see Section 3.2). Ultimately, the most energy-efficient implementation of an application is one that has dedicated hardware suited for the application and context, a solution, i.e., often economically infeasible. More generalized hardware leads to lower cost, but typically greater energy consumption.

The most significant factor in improving the energy efficiency of software is the developer. Developers can adjust their software to be more efficient on their chosen platform; however, this requires detailed understanding and creativity. It is currently impossible for automated processes to fully customize an algorithm to take full advantage of available hardware features, and thus, it is up to the creativity of the developer to do so. To create the best opportunities for efficiency, the developer should be provided with as much information about their software's energy consumption as possible: at all levels, from the switching cost to static power and system-level energy consumption. Enabling the developer in such a way will allow for energy-aware exploration of the design space, allowing informed decisions to be made so that the developer can pursue minimal energy. Such benefits are not limited to individual applications either: developers of software libraries, compiler optimizations, code schedulers, and any other software facility will be able to make energy-orientated design decisions. Better software design for hardware should also enable better selection of hardware. Understanding precisely how a piece of software consumes energy would allow more informed selection of heterogeneous systems, particularly if the heterogeneous hardware's features could be characterized in a way that can be matched against software features. Such a process would greatly simplify the benchmarking and evaluation required when selecting a platform to develop on.

In summary, new paradigms and tools for the design of software, based on energy transparency, are required if the energy consumption is to be reduced. Currently, the main drivers for software production are high performance, minimizing time for operation, and minimizing production cost. These drivers must include minimizing energy. For this to succeed, designers should be able to:

- Allocate compute resources in units of energy, not just time.
- Capture both execution duration and power efficiency.
- Force application developers to think about energy and understand and energy model of the software.
- Access energy-aware tools for developing energy-efficient software and code.

Tools such as energy modelling of software, software energy analysis methods, energy profiling, and metrics for numerics required for the development of energy-aware software and algorithms and of energy transparency are discussed in Chapters 5 and 6.

3.4. Communications

The ability to communicate information among devices, memory, storage, and systems is fundamental to ICT systems. While many concentrate on the energy from switching logic and memory devices, communication between devices and especially circuits or systems can be many orders

of magnitude greater than logic operations. Efforts to reduce data transfer (and data reduction/compression methods) at the architecture and software level have been alluded in previous sections. Here, we summarize the different sources of the energy overhead of communications.

3.4.1. Chip-level: interconnects and electrical wires

Interconnect scaling has a significant impact on microarchitecture as it governs the speed at which data can be moved around a chip. In a micro-architecture with many cores, each with multiple functional blocks, the interconnect can become a bottleneck to performance. While transistors have benefited from performance increases as they have shrunk, interconnects have not, their delay product remaining close to constant [38]. The interconnect now plays a large part in the delays present in memory hierarchies. Thus, the structure of cache hierarchies—something, i.e., very much a micro-architectural choice—is governed not just by the speed of the storage cells, but the delays in moving data between those caches and the processor. Fast, first-level caches cannot be large, because it becomes impossible to transport data across them and still close timing constraints. Thus, modern high-performance processors have multiple cache layers, increasing in capacity and latency as they get further from the computational part of the processor.

For any electrical wire, the energy consumed to transmit a bit of information is related to the capacitance, C and the voltage, V. If the capacitance is approximated by $C \sim \varepsilon_0 L$ assuming an isolated wire where ε_0 is the permittivity of a vacuum, and L is the length, then the energy for transmitting a bit of information in a metal interconnect or wire can be approximated as

$$E_{\text{wire}} = \frac{C V^2}{2} \sim \frac{\varepsilon_0 L}{2} \left(\frac{k_B T}{e} \right)^2 \quad (1)$$

This gives a fundamental limit of 3×10^{-15} J per bit.m for isolated interconnects or wires [39]. The other fundamental limit is shot noise limit relating to the quantum of electric charge in electrons being transported along interconnects. At 300 K this corresponds to energy of 2.9×10^{-21} J/bit. In real systems, the actual values are significantly larger than the fundamental limits. In particular, the signal to noise and fan-out require significantly higher energy consumption especially when circuit speeds are high.

3.4.2. System-level: photonics and wireless

At the moment optical cables between boards and racks are available for Data Centres and HPC, but there is still significant copper interconnects before getting to the microprocessor, on-chip memory, or disk storage. The development of integrated Si photonics where the enormous yields of silicon foundries can be used to produce far cheaper and larger bandwidth (through parallel channels) has the potential to reduce energy per bit far faster than older technologies. Such technology is being developed for not only chip-to-chip photonic communications but also on-chip communications for the higher level and longer interconnects. If the volumes can be reached, then the costs should allow large-scale deployment. Miller [40] has investigated the requirements for chip-to-chip and on-chip photonic communications that provide an idea of the ultimate limits for photonic communications. The limits on individual photonic components is likely to be around 10 fJ/bit for the most power

http://dx.doi.org/10.5772/66678

hungry, suggesting that the limit for short-distance photonic communications (<10 m distance) is around 100 fJ/bit.

For the rack-to-rack optical interconnects for exascale computing for 2019, the US Department of Energy calculated that every pJ/bit of optical power results in a total contribution of 0.8 MW for the complete system power [41]. As of 2014, a typical rack-to-rack optical interconnect at 40 Gbits/s is operating at around 40 pJ/bit. A number of authors have estimated the energy cost of sending information over the Internet [42]. For example in 2009, Baliga et al. [43] undertook a modelling study that included the energy consumption of the core, metro and edge, access, and video distribution networks. This included the energy consumption from switching and transmission equipment. They found energy consumption per bit of 75 μJ at low access rates that decreased from 2 to 4 μJ at an access rate of 100 MB/s. This study estimated that the Internet communication components alone accounted for 0.4% of electricity consumption in broadband-enabled countries at that point, with this percentage predicted to increase significantly as the bandwidth increases. They modelled how the energy per bit will reduce depending on how fast photonic technology improves and with a 10% rate of improved energy efficiency provides 700 nJ/bit by 2023, while a 20% rate of improvement results in 120 nJ/bit. The biggest issue is whether these reductions are aggressive enough to counteract the increase in the bandwidth of Internet and microwave wireless traffic. A key finding of the study was that the predicted rate of improvement with the future photonic technology in reducing the energy consumption per bit was not sufficient to reduce the energy consumption in the future as demand increases.

The IoT will require increased communication bandwidth, but as many of the systems are battery-powered and require wireless transmission, there are significant incentives to minimize communication bandwidth, distance, and time to save battery power. High-definition video is another issue. As use of high-definition video on demand increases, the systems architecture of the Internet to reduce long-haul delivery requires investigation with an aim to reduce energy consumption which is proportional to the bandwidth. For wireless communications the energy to transmit a bit of information is given by

$$E_{\text{wireless}} = N_{\text{photons}} E_{\text{photons}} \tag{2}$$

where N_{photons} is the number of photons given for uniformly radiating wireless as $N_{\text{photons}} \sim \frac{4\pi r^2}{\lambda^2}$ and E_{photon} is the photon energy given by $E_{\text{photons}} = h\nu = \frac{hc}{\lambda}$. The energy to transmit a bit of information is therefore given by [39]

$$E_{\text{wireless}} \sim \frac{4\pi r^2 hc}{\lambda} \tag{3}$$

Figure 10 presents the ultimate energy per bit that can be transmitted over 10 m distance by wireless using the single photon limit calculation and compares this to different wireless technologies. All the wireless technologies are around the 60—200 nJ range per bit, which is many orders of magnitude above the fundamental limits indicating that there is significant potential for improvements in the energy consumption of wireless communications. Chapter 9 of Vol. 1 presents an introduction to power consumption in wireless sensor networks including power consumption assessment via modelling and measurements. The evolution and state-of-the-art of wirelessly communicating networks of embedded computers and their energy efficiency are described in Chapter 7.

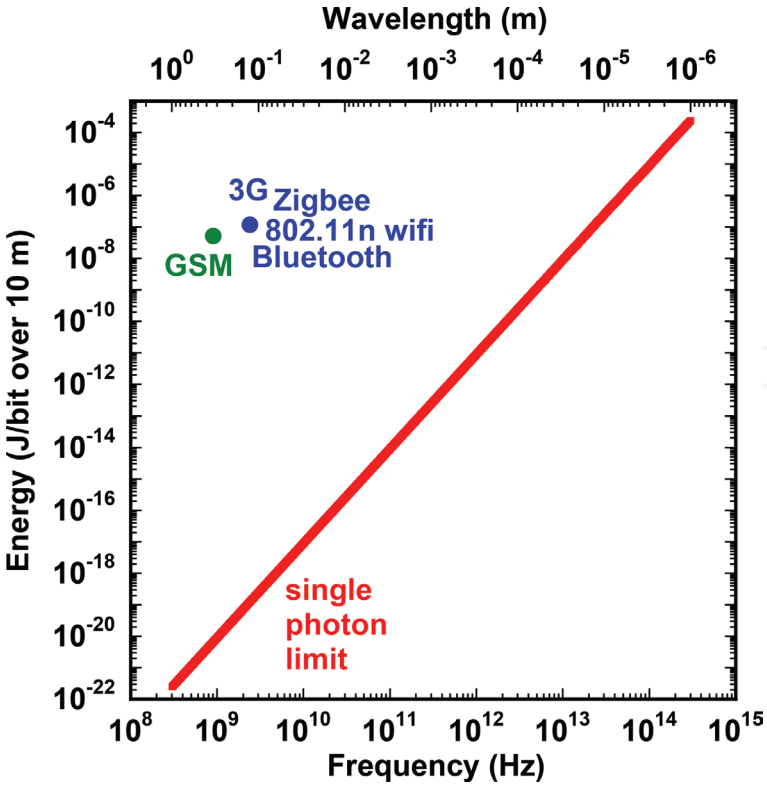

Figure 10. The energy per bit to transmit wireless data 10 m using a number of technologies at a range of transmission frequencies as compared to the fundamental single photon limit for omni-directional wireless transmission.

3.5. Energy sources and power management

Already many HPC cloud servers are being located so that renewable energy can be used to power at least part of the systems. Some of the best examples of low carbon renewables that can be used are photovoltaic, hydro, and wind (see **Figure 6**). In every case, the correct environment is required for each of the renewable technologies. A good example of this is photovoltaics that is being heavily used for a significant number of Data Centres for cloud computing. **Figure 11** provides the available energy averaged over 24 hours and 365 days of the year for a range of cities around Earth. The actual available energy over 24 hours will be these values times the conversion efficiency for the PV technology being used. The record PV efficiencies can be found at the charts published by the National Renewable Energy Laboratory (NREL) [http://www.nrel.gov/ncpv/images/efficiency_chart.jpg], but the majority of deployed PV solar farms use crystalline Si PV cells typically with starting efficiencies of 20–22%. Therefore, a solar farm in Athens will produce about 41 W/m^2 and to power a 10 MW HPC or cloud Data Centre will require at least 250,000 m^2 of PV area. Due to dust and dirt along with the harsh

http://dx.doi.org/10.5772/66678

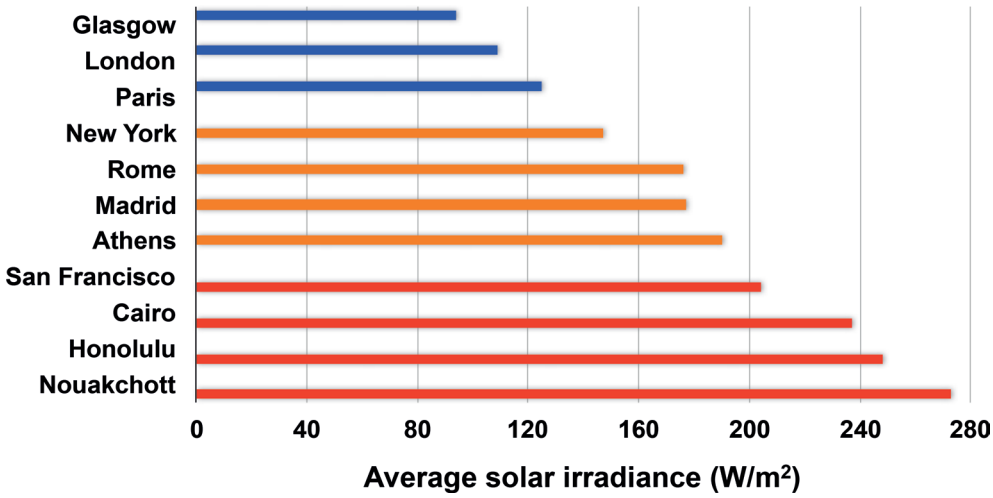

Figure 11. The average solar power available at different points on the earth integrated over 24 hours and 365 days of the year. The peak power is significantly higher than these values, but these are the ones available 24/7 (Source [44]).

environment that PV operates in, the efficiency drops off with time so a significantly larger area is required for long-term sustainable energy generation. Also, as mentioned at the end of Section 2, a significant challenge is that substantial energy storage is required to capture the energy that is only available during the day so it can be deployed at night.

The control and conversion of electric power through efficient power electronics is another issue to consider for power management on a large scale. While national grids that transmit electricity to industry and domestic properties for use all transmit using high-voltage AC to minimize losses, all ICT systems operate with DC power supplies requiring power transformation and management. At present switch-mode power supplies dominate AC to DC conversion for most ICT devices such as PCs, laptops, smart phones, and mobile phones. Such converters are used since they have greater efficiency than other technologies such as linear power supplies because the switching transistor dissipates little power when acting as a switch and spends very little time in any high-dissipation energy transition. Silicon power switches include power MOSFETs and insulated gate bipolar transistors (IGBT) depending on the power and regulation requirements. To improve efficiencies in power conversion, most research is concentrating on developing new materials for power switches which reduce the ohmic losses and the on-resistance. SiC and GaN are the main materials being developed as the wider bandgap and higher electrical conductivities should improve the conversion efficiencies. As an example, GaN is predicted to have 50% higher conversion efficiency than Si for power switches and just converting all electrical drives to GaN is predicted to save 9% of the electricity consumption in the UK that corresponds to removing the equivalent generation capacity of five advanced gas cooler nuclear reactors (UK Department for Business, Innovation and Skills October [45]). The potential for energy savings across Europe and the world are enormous and have been predicted to be of the order of €1400 Bn per annum (UK Department for Business, Innovation and Skills October [45]) if GaN technology fully replaces

present Si power switches. An area of primary interest is complete integration of both modular and granular electronic power converters on-die and within package (power supply-on-chip) [46]. Following the trends of More Moore and More than Moore, this integration would deliver ever-greater current density, voltage regulation, and optimized control, form factor reduction, high efficiency, and cost reduction to meet performance requirements of emerging ICT systems.

Energy and power management in smart autonomous sensors needs to be embedded within the system architecture utilizing energy harvesting from the environment, energy storage devices, and efficient power distribution architecture. Similar to renewable energy, the energy sources must be chosen dependent on the operation environment. Energy harvesting mechanisms typically convert ambient kinetic energy, wasted energy (heat), and electromagnetic radiation (light, RF) into useful electricity and the challenge is to maximize the available extraction and conversion efficiency. Vibrational energy harvesters around the size of a drink are able to extract up to 5 mW of power from kinetic energy in pumps, trains, or vehicles. Such systems are now deployed in many industrial and transport applications. Thermoelectrics can be used to convert waste heat into electricity. Thermoelectric and photovoltaic (PV) energy conversion seems to be most promising for a large range of applications for small-to-medium power generation (and large for PV solar farms; see **Figure 6**). A temperature gradient is required to generate electricity with the voltage and power output dependent on the size of this temperature gradient. Such systems are used for smart thermostat controls of heaters and heating systems and also are being developed for industrial applications and automotive (to improve fuel consumption of vehicles). For autonomous sensors which can use PV, it is a reliable source during the day and batteries or super capacitors can be used to store up for nighttime use. Even in northern EU countries, the generation levels are around 20 W/m^2 per day, providing significant energy for most autonomous sensors. Indoor PV has significantly lower energy available typically 100 times lower than outdoor direct sunlight. Also most indoor available light is diffuse, scattered off different surfaces, and so capturing this light is far more difficult than capturing direct sunlight. The efficiency to capture diffused sunlight results in PV cells with efficiencies only up to about 10% at present. This is still useful for many autonomous sensors. A comprehensive presentation of energy harvesters can be found in Vol. 1. The physics of thermoelectrics and PV is presented in Chapter 9 for completeness.

As for HPC/Data Centres, energy storage and power management present a significant challenge for smart autonomous sensors too. There is a need to bridge the gap between the energy requirements for the ICT system operation and the energy supply from harvesting sources and storage to enable true autonomy (incl. wireless communication). In addition to decreasing the energy demand from the electronics and increasing the efficiency of the energy harvesters, an increase in storage capacity of batteries and the efficiency of power management electronics is required. The ideal scenario is an ICT system that integrates energy storage with energy harvesting components to provide power on demand to the electronic sensing, communication and display components, and it operates over the anticipated device lifetime of years. Ultimately, the energy supply components, harvesting and storage should occupy a footprint on chip no larger than the electronics it drives, with 1 mm^2 as an attractive long-term target for both. No such energy storage component exists today.

http://dx.doi.org/10.5772/66678

Batteries are the most common energy storage option, and since their introduction in the 1990s, lithium-ion batteries have exhibited the highest energy density that has been gradually improved in the intervening period from ~200 Wh/l to ~700 Wh/l through the use of improved materials and processing (see Chapter 6 of Vol 1 for a comprehensive overview of battery materials and architectures). Solid-state microbatteries that can be processed/integrated on silicon substrate exhibit similar volumetric energy densities with micron scale thin film materials. This necessitates a large area format consistent with the large areas or volumes required by solar or vibrational harvesting, respectively. Solid-state devices do offer capacity retention over thousands of cycles [47] that matches the device lifetime requirements. In the 2D, thin-film geometry, current deposition techniques, and lithium-ion diffusion characteristics in the solid-state limit the electrode thickness to several micrometers resulting in a battery dominated by the substrate and other inactive cell components. As thin films, these 2D formats typically exhibit energy densities of ~6 Wh/mm^2 or 0.2 J/mm^2. An energy budget of 1 mWh/day can support a wireless sensor node (WSN) used in building energy management with sensing and transmission every 20 minutes [46]. Clearly, significant advances are required for energy storage devices to meet the demand in a reasonable footprint. The key challenges are to realize improved energy storage in a significantly decreased footprint for ICT integration and high rate (power) capability during device interrogation and to decrease recharge time. These challenges require:

- Higher energy density materials, particularly at the cathode, where the current material, $LiCoO_2$, is 25 times less energy dense than the lithium metal anode.
- 3D or 1D active materials structuring with increased aspect ratio providing additional material (stored energy) with respect to planar commercial thin film microbatteries.
- Nanoscale active materials with improved electronic conductivity or core/shell structures [48] to facilitate high rate solid-state lithium ion transport.

Furthermore, current solid-state microbatteries cannot meet the needs of the ICT systems at peak power during measurement and transceiver operation and require a hybrid energy source with a supercapacitor. This is due to the lithium diffusion limitation in the solid-state electrolyte and cathode. On the other hand, the solid-state construction does facilitate the use of lithium metal anodes that have a large energy capacity (3600 mAh/g) in comparison with the typical carbon anodes (372 mAh/g) of most lithium-ion batteries. If nonsolid-state electrolytes are to be utilized for higher power outputs, then alternative high-energy intercalation anodes such as Sn (990 mAh/g), Ge (1600 mAh/g), or Si (4200 mAh/g) will be required to prevent dendritic short circuits on cycling. Core/shell [49] versions of these anodes may be required to alleviate mechanical stresses, leading to poor cycling behaviour and improve the electronic conductivity to access all of the high aspect ratio structures.

Disruptive battery technologies such as a Li/sulphur or Li/air can achieve the energy storage requirements of the ICT community. A theoretical energy density of 2.8 mWh/mm^3 (10 J/mm^3) has been estimated for a Li-air battery [50] with nonaqueous electrolytes. It is recognized that the Li/air and other high energy systems have to overcome many obstacles before they will

be in widespread deployment, but many researchers predict that this could be over the next 10—15 years. A very challenging scaling requirement is shown in **Figure 12** and that must be coupled with decreased power requirements in the ultimate device. It is clear from the values in **Figure 12** that both new materials and 3D structuring of the materials are required to enable the decreased footprint desired for autonomous systems. An increasing focus on improved nanoarchitectures, electrode materials, and integrated current collectors is required to surmount these obstacles and deliver high-energy density solutions to meet the needs of the electronics industry. On the other hand, supercapacitors are emerging as an attractive candidate to complement advanced lithium batteries [52]. The key advantages of supercapacitors for such applications include high-power density, faster charge and discharge rates, and improved cycle-life.

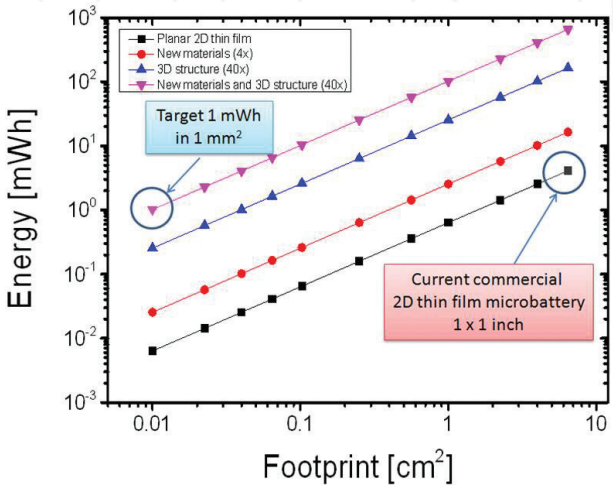

Figure 12. A roadmap for microbattery energy storage requiring the development and integration of new materials that could yield up to four times improvement in stored energy and micro or nanoarchitectures to increase the material quantity and surface area to deliver 1 mWh of energy in a 1 mm² footprint (Source [51]).

While traditional electrochemical supercapacitors can provide very high specific energy, preventing leakage of the supercapacitors liquid electrolyte solution in a portable or implantable device may be challenging. Also, since exposure of the liquid electrolyte solution to moisture in air can adversely affect its performance, special manufacturing needs are required that to date have prevented integration of supercapacitors into standard manufacturing processes. Thus, there is a market pull for a mechanically durable high-performance solid-state supercapacitor that can be fabricated using standard semiconductor manufacturing processes. While some solid-state supercapacitors (SSCs) with highly desirable power and energy density attributes have been demonstrated, significant practical challenges still remain in order to deliver

http://dx.doi.org/10.5772/66678

high-performance products to meet the future demands of ICT applications. Desired targets for future solid-state supercapacitors to meet these demands would include:

- High energy density >10 Wh/kg.[1]
- High power density >1 × 10^6 W/kg.
- Wide operating temperature window (−20°C to +70°C).
- Low equivalent series resistance (ESR) <10 mΩ.
- Long cycle life >1 × 10^6 cycles.

4. Conclusion

The key message that has become clear from investigating each of the levels of the system stack is that overall energy efficiency can be optimized more aggressively when the design at one level understands the energy issues at another level. This is especially true for levels that control the behaviour of another level, e.g., circuit architecture on devices or software on circuit architecture. At the start of ICT systems in the 1960s and 1970s, it was possible for engineers to understand all levels of the system stack and design accordingly. As each level has become more complex and the number of transistors and lines of code have moved from thousands into many billions, it has become more difficult for people to understand all levels of the system. A clear message is that only through the joint understanding of how different levels of the stack must be designed to reduce energy consumption will optimum ICT solutions that minimize the use of energy be found. As communication is the biggest energy consumption, and sustainable energy is environmental dependent, what is the optimum distribution and location of servers for the cloud if it is driven by sustainable energy ICT? Similarly, as the energy per bit of wireless communication scales with distance, what is the optimum distribution of smart autonomous sensors and data exchange to minimize energy consumption? It should be pointed out that in some circumstances minimizing communication is not the lowest energy solution—e.g., the processing required for heavy compression can cost more energy than you save under most transmission scenarios. If significant energy reductions are to be achieved, suitable education and tools are essential where an expert (e.g., in software) is provided with sufficient knowledge to understand the energy impact (of code) at the other levels of the stack. Realizing this vision requires interdisciplinary research at the boundaries of multiple scientific domains (materials science, physics, electrical engineering, software engineering, and mathematics), as well as developing and integrating innovations in several research areas, e.g., materials modelling and fabrication, device and computer engineering, cooling design, large-scale computing system simulation, software generation and optimization, statistical network modelling, and model predictive control theory.

[1] The energy density of current commercial off-the-shelf supercapacitors is relatively low at less than ~ 10 Wh/kg [53] with current solid state supercapacitor equivalents at ~ 1 Wh/kg [54].

Acknowledgements

This chapter has been broadly based on the Strategic Research Agenda (SRA) of the ICT-Energy project funded by the European Union (Ref: 611004). Many people have contributed in developing this SRA: Giovanni Ansaloni (EPFL), David Atienza (EPFL), Micheal Burke (Tyndall National Institute), Zbigniew Chamski (MpicoSys), Adrian Cristal (Barcelona Supercomputer Centre), Kerstin Eder (University of Bristol), Pablo Garcia (EPFL), Vincent Heuveline (Heidelberg University), Steve Kerrison (University of Bristol), Louise Krug (BT), Andrew Lord (BT), Jeremy Morse (University of Bristol), Simeon Oxizidis (International Energy Research Centre at Tyndall) James Rohan (Tyndall National Institute), Martin Wlotzka (Heidelberg University, M. Fernando Gonzalez Zalba (Hitachi Cambridge Laboratory), Olivier Zendra (Inria), Victor Zhirnov (Semiconductor Research Corporation). Contributions were also taken from the ICT Energy workshop "Future Energy" in the ICT Research Agenda on the 15 September 2015 in Bristol, where this SRA was presented, and a number of working groups provided feedback that has been included. Not all the names of the participants have been included in the above list. The participants included a wide range of academics through the system stack and companies including ARM and Intel.

Author details

Giorgos Fagas[1,*], John P. Gallagher[2], Luca Gammaitoni[3] and Douglas J. Paul[4]

*Address all correspondence to: georgios.fagas@tyndall.ie

1 Tyndall National Institute, UCC, Cork, Ireland

2 IMT, Roskilde University, Roskilde, Denmark

3 NiPS Laboratory, Department of Physics & Geology, University of Perugia, Perugia, Italy

4 School of Engineering, University of Glasgow, Glasgow, UK

References

[1] S.G. Anders Andrae and T. Edler, "On Global Electricity Usage of Communication Technology: Trends to 2030," Challenges 6, 117–157 (2015).

[2] United Nation's Brundtland Commission "Our Common Future" (1987)—Available at link: http://www.un-documents.net/our-common-future.pdf

[3] G. Fagas, L. Gammaitoni, D. Paul and G. Abadal Berini (eds.), ICT—Energy—Concepts Towards Zero—Power Information and Communication Technology," InTech ISBN 978-953-51-1218-1, DOI: 10.5772/55410 (2014)–Available at link: http://www.intechopen.com/books/ict-energy-concepts-towards-zero-power-information-and-communication-technology

[4] http://www.top500.org.

[5] R.H. Dennard et al., "Design of Ion Implanted MOSFET's with very Small Physical Dimensions," IEEE Journal of Solid-State Circuits 9(5), pp. 256–268 (1974).

[6] F. Pollack "New Microachitecture Challenges in the Coming Generations of CMOS Process Technology," 32nd International Symposium on Microarchitecture (Micro32) (1999)—Available at link: http://research.ac.upc.edu/HPCseminar/SEM9900/Pollack1.pdf

[7] International Atomic Energy Agency, "Nuclear Power Plant Design Characteristics" (2007)—Available at link: http://www-pub.iaea.org/mtcd/publications/pdf/te_1544_web.pdf

[8] ARM White Paper, "big.LITTLE Technology: The Future of Mobile Making very high performance available in a mobile envelope without sacrificing energy efficiency" (2013) —Available at link https://www.arm.com/files/pdf/big_LITTLE_Technology_the_Futue_of_Mobile.pdf

[9] http://www.gartner.com/technology/research.jsp

[10] United Nations Department of Economic and Social Affairs "World Population Prospects: The 2012 Revision" (2012)—Available at link: http://esa.un.org/wpp/unpp/panel_population.htm

[11] Cisco White Paper "The Zettabyte Era: Trends and Analysis" (2015)—Available at link: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html

[12] M. Pickavet, W. Vereecken, S. Demeyer, P. Audenaert, B. Vermeulen, C. Develder, D. Colle, B. Dhoedt and P. Demeester, "Worldwide Energy Needs for ICT: The Rise of Power-Aware Networking," 2nd International Conference on Advanced Networks and Telecommunication Systems, 2008, pp. 1–3 (2008).

[13] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet and P. Demeester, "Trends in Worldwide ICT Electricity Consumption from 2007 to 2012," Computer Communications 50, pp 64–76 (2014).

[14] J. Malmodin, A. Moberg, D. Lundén, G. Finnveden and N. Lövehagen, "Greenhouse Gas Emissions and Operational Electricity Use in the ICT and Entertainment & Media Sectors," Journal of Industrial Ecology 14, pp. 770–790 (2010).

[15] Global e-Sustainability Initiative "SMARTer2020: The Role of ICT in Driving a Sustainable Future" (2013)—Available at link: http://gesi.org/portfolio/report/72

[16] International Energy Authority, "Key World Energy Statistics" (2014)—Available at link: http://www.iea.org/statistics/

[17] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," IBM Journal of Research Development 5, 183–191 (1961).

[18] V.V. Zhirnov, R.K. Cavin III, J.A. Hutchby and G.I. Bourianoff, "Limits to Binary Logic Switch Scaling—A Gedanken Model," Proceedings of the IEEE 91, pp. 1934–1939 (2003).

[19] International Technology Roadmap for Semiconductors—Available at link: http://www.itrs.net/

[20] D.E. Nikonov and I.A. Young, "Uniform Methodology for Benchmarking beyond-CMOS Logic Devices," Proceedings IEDM 12, pp. 576–672 (2012).

[21] M. López-Suárez, I. Neri and L. Gammaitoni, "Sub-kBT Micro Electromechanical Irreversible Logic Gate," Nature Communications 7, 12068 (2016).

[22] B.K.G. Brill, "The Invisible Crisis in the Data Center: The Economic Meltdown of Moore's Law," White Paper, Uptime Institute, Rev. 2, 2007–12—Available at link: http://www.mm4m.net/library/The_Invisible_Crisis_in_the_Data_Center.pdf

[23] H. Mamaghanian, N. Khaled, D. Atienza Alonso and P. Vandergheynst, "Design and Exploration of Low-Power Analog to Information Conversion Based on Compressed Sensing," IEEE Journal of Emerging and Selected Topics in Circuits and Systems 2(3), pp. 493–501, (2012).

[24] S. Benatti, B. Milosevic, F. Casamassima, P. Schonle, P. Bunjaku, S. Fateh, Q. Huang, L. Benini, "EMG-based Hand Gesture Recognition With Flexible Analog Front End," in Proceedings of International IEEE Conference on Biomedical Circuit and Systems (BIOCAS 2014), Lausanne, Oct 2014

[25] N. Planes et al., "28 nm FDSOI Technology Platform for High-Speed Low-Voltage Digital Applications," Proceedings of Symposium VLSI (2012).

[26] J.-P. Colinge and J. Greer, "Nanowire Transistors," Cambridge University Press (Cambridge, UK), ISBN: 978-1107052406 (2016).

[27] A. Sridhar, M.M. Sabry, P. Ruch, D. Atienza, B. Michel, "PowerCool: Simulation of Integrated Microfluidic Power Generation in Bright Silicon MPSoCs," oral, ICCAD, November 2014, San Jose, CA, USA.

[28] N. Rasmussen, "Guidelines for Specification of Data Center Power Density," (2005)—Available at link: http://www.apcdistributors.com/white-papers/Cooling/WP%20120%20Guidelines%20for%20Specification%20of%20Data%20Center%20Power%20Density.pdf

[29] V. Hanumaiah and S. Vrudhula, "Energy-efficient Operation of Multicore Processors by DVFS, Task Migration and Active Cooling," IEEE Transactions on Computers 63(2), pp. 349–360, (2012).

[30] P. Bassett and M. Saint-Laurent, "Energy Efficient Design Techniques for a Digital Signal Processor," 2012 IEEE International Conference on IC Design Technology (ICICDT), pp. 1–4 (2012).

[31] ARM Cortex-A15 MPCore Processor Technical Reference Manual, ARM, pp. 53–63 (2013).

[32] R. Bahar and S. Manne, "Power and Energy Reduction via Pipeline Balancing," Proceedings 28th Annual International Symposium on Computer Architecture, 2001, pp. 218–229 (2001).

[33] H. Zeng, J. Wang, G. Zhang, and W. Hu, "An Interconnect-Aware Power Efficient Cache Coherence Protocol for CMPs," IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2008 pp. 1–11 (2008).

[34] A.Z. Jooya and M. Analoui, "Program Phase Detection in Heterogeneous Multi-Core Processors," 14th International CSI Computer Conference, CSICC 2009 pp. 219–224 (2009).

[35] K. Changkyu, S. Sethumadhavan, M. S. Govindan, N. Ranganathan, D. Gulati, D. Burger, and S. Keckler, "Composable Lightweight Processors," 40th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2007 pp. 381–394 (2007).

[36] Z. Rakossy, T. Naphade, and A. Chattopadhyay, "Design and Analysis of Layered Coarse-Grained Reconfigurable Architecture," 2012 International Conference on Reconfigurable Computing and FPGAs (ReConFig) 2012 pp. 1–6 (2012).

[37] K. Roy and M.C. Johnson, "Software Design for Low Power," in "Low Power Design in Deep Submicron Electronics," W. Nebel and J. Mermet (Eds.), Kluwer Nato Advanced Science Institutes Series, Vol. 337. Kluwer Academic Publishers, Norwell, MA, USA, pp 433–460 (1997).

[38] M. Bohr, "A 30 Year Retrospective on Dennard's MOSFET Scaling Paper," IEEE Solid-State Circuits Society, Newsletter 12(1), pp. 11–13 (2007).

[39] V.V. Zhirnov, "Fundamentals of Energy Consumption in ICT Devices," NIPS/ICT Energy Summer School (2014)–Available at link: http://www.nipslab.org/sites/nipslab.org/files/NiPS2014_Zhirnov_en%20cons%20ICT.pdf

[40] D.A.B. Miller, "Device Requirements for Optical Interconnects to Silicon Chip," Proceedings of IEEE 97(7), pp. 1166–1185 (2009).

[41] http://www.hoti.org/hoti20/slides/Fuad_Doany_IBM.pdf

[42] J. Baliga et al., "Green Cloud Computing: Balancing Energy in Processing, Storage and Transport," Proceedings of IEEE 99(1), pp. 149–167 (2011).

[43] J. Baliga et al., "Energy Consumption in Optical IP Networks," Journal of Lightwave Technology 27(13), pp. 2391–2403 (2009).

[44] G. Boyle, "Renewable Energy: Power for a Sustainable Future," Oxford University Press (Oxford, UK), ISBN: 978-0199545339. (2012).

[45] UK Department for Business, Innovation and Skills, "Power Electronics: A Strategy for Success," October (2011)–Available at link: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31795/11-1073-power-electronics-strategy-for-success.pdf

[46] C.Ó. Mathúna, T. O'Donnell, R. Martinez, J.F. Rohan and B. O'Flynn, "Energy Scavenging for Long-Term Deployable Mote Networks," Talanta 75, 613–623 (2008).

[47] N.J. Dudney, "Solid-state thin-film rechargeable batteries ", Journal of Materials Science and Engineering B-Solid State Materials for Advanced Technology 2005, 116, 245.

[48] M. Hasan, T. Chowdhury, J.F. Rohan, "Nanotubes of Core/Shell Cu/Cu2O as Anode Materials for Li-ion Rechargeable Batteries ", Journal of the Electrochemical Society A682, 157(2010).

[49] L.F. Cui, R. Ruffo, C.K. Chan, H.L. Peng, Y. Cui, "Crystalline-Amorphous Core-Shell Silicon Nanowires for High Capacity and High Current Battery Electrodes", Nano Letters 9, 491 (2009).

[50] J.P. Zheng, R.Y. Liang, M. Hendrickson, E.J. Plichta, "Theoretical energy density of Li–air batteries", Journal of the Electrochemical Society 155, A432 (2008).

[51] W. Wang, J.F. Rohan, N. Wang, M. Hayes, A. Romani, E. Macrelli, M. Dini, M. Filippi, M. Tartagni and D. Flandre, Chapter 9—"Smart Energy Management and Conversion in Beyond CMOS Nanodevices," 1 (2014) Editor F. Balestra, Wiley, pp 249–276, ISBN: 978-1-84821-654-9.

[52] F. Gonzalez and P. Harrop, "Batteries & Supercapacitors in Consumer Electronics 2013–2023: Forecasts, Opportunities, Innovation," IDTechEx (Cambridge, UK) (2014).

[53] A. Burke, Z. Liu, H. Zhao, "Present and future applications of supercapacitors in electric and hybrid vehicles," IEEE International Electric Vehicle Conference, DOI: 10.1109/IEVC.2014.7056094 (2014)

[54] P. Banerjee, I. Perez, L. Henn-Lecordier, S. B. Lee, and G.W. Rublof, "Nanotubular metal-insulator-metal capacitor arrays for energy storage" Nature Nanotechnology 4, 292 (2009)